

Oponentní posudek na magisterskou práci Daniela Bicáka

EVALUACE GEOGRAFICKÉHO RANDOM FOREST ALGORITMU V POSOUZENÍ SUCHA

Oponovaná práce Daniela Bicáka s cílem evaluaovat geografický Random Forest při hodnocení zemědělského sucha splňuje podmínku náročnosti kladenou na diplomovou práci a má kartograficko-geografický charakter.

Formulace cílů

Student si stanovil jako cíl evaluaovat rozšíření široce používaného algoritmu strojového učení Random Forest na geografický Random Forest (GRF), což je úloha plně zapadající do oboru geoinformatika. Toto rozšíření pak student aplikoval při hodnocení zemědělského sucha. Dílčí cíle jsou:

- evaluace geografického rozšíření algoritmu RF pro predikci zemědělského sucha, včetně ladění parametrů modelu,
- ověření lokálních a globálních faktorů zemědělského sucha,
- a tvorba mapy prostorové variability zranitelnosti vůči suchu v České republice s využitím navrženého a parametrizovaného modelu.

V úvodu diplomové práce jsou zároveň popsány a vymezeny výzkumné otázky úlohy.

Práce s literárními zdroji

Rozsah literárních zdrojů použitých v řešení práce je velký, překračuje 100 zdrojů. Student studoval zadanou problematiku do hloubky. Teoretický rozbor zahrnuje popis problematiky zemědělského sucha, indikátorů zranitelnosti, dále strojového učení obecně, algoritmu Random Forest a možností jeho prostorového rozšíření včetně parametrizace takového modelu.

Vhodnost postupu a technologie zpracování

Kapitola Metodika je vhodně strukturovaná, popisuje tři základní fáze řešení práce. Dílčí podkapitoly vymezují metodický návrh vlastního řešení. Následuje popis vstupních dat a technologických komponent programovacího jazyka Python.

Úroveň zpracování a metodické splnění cílů

Autor vytvořil vlastní implementaci algoritmu geografický Random Forest a aplikoval tento model v několika úlohách k naplnění cílů. Součástí dokumentovaného řešení je průzkum vstupních dat, ověření kolinearit vstupních příznaků, prostorové vzorkování rozsáhlého data setu, parametrizace modelů, včetně analýzy vlivu jednotlivých parametrů na kvalitativní indikátory modelu až po analýzu hyperparametrů geografického rozšíření RF. Součástí práce je i analýza prostorových závislostí hyperparametrů GRF.

Kapitola 6.3 přináší souhrn výsledků výzkumu a porovnává jednotlivé varianty modelů, standardní Random Forest s rozšířením o prostorové souřadnice jako vstupní příznaky modelu, geografický Random Forest a lokální GRF model. Výsledky jsou diskutovány a porovnány s existujícími pracemi publikovanými v IF časopisech. Nedílnou kapitolou je rozbor důležitosti příznaků v jednotlivých variantách modelu, který umožňuje interpretovat vliv lokálních a globálních faktorů na sledovaný jev.

Práci bych doporučil rozšířit o úryvek kódu v jazyce Python, případně pseudo-kódu, popisující algoritmus GRF a jeho parametry podobně jako uvádí autor na str. 23 obecný algoritmus Random Forest (dle Hastie et al., 2009).

Vlastní přínos autora vidím především ve vlastní programové implementaci GRF a jeho demonstraci včetně interpretace poukazující na důležitost prostorového kontextu v analýzách geografických dat. Diplomová práce zároveň otevírá řadu výzkumných otázek souvisejících s využitím algoritmů strojového učení v geovědách.

Diplomová práce splňuje ostatní formální náležitosti, práce je bohatě ilustrována obrázky a doplněna tabulkami. Předloženou a oponovanou magisterskou práci Daniela Bicáka doporučuji k obhajobě a navrhuji klasifikovat výborně.

V Praze dne 13. září 2021

Ing. Lukáš Brodský, PhD.

Katedra aplikované geoinformatiky a kartografie

Přírodovědecká fakulta Univerzity Karlovy